scottish music review

A New Approach to Onset Detection: Towards an Empirical Grounding of Theoretical and Speculative Ideologies of Musical Performance

Keziah Milligan Nick Bailey Graham Hair

1: A summary of the current empirical study

This article assesses aspects of the current state of a project which aims, with the help of computers and computer software, to segment soundfiles of vocal melodies into their component notes, identifying precisely when the onset of each note occurs, and then tracking the pitch trajectory of each note, especially in melodies employing a variety of non-standard temperaments, in which musical intervals smaller than 100 cents are ubiquitous. From there, we may proceed further, to describe many other "micro-features" of each of the notes, but for now our focus is on the onset times and pitch trajectories.

Stated in such bald terms, such a description makes the current project sound like a quite small and modest one, but in fact, the very first hurdle, automated onset detection, presents a quite difficult challenge, and attempts to solve it have thus far met with only patchy success. The project's larger aim is to facilitate the efficient advance of one particular approach to the study of performance, using objective contemporary machine-based methods.

To be sure, our study considers the problems involved, but not just problems in getting correct answers. In what follows we aim also to take account of the fact that the study of music is always cast in several different cultural contexts: contexts such as widely-accepted principles of analysis, the capacity for any new work to challenge the truth, exactitude or appropriateness of those principles, and the influence on any empirical work in musicology of traditions taken from other disciplines and sub-disciplines. Very often the evidence from studies of particular phenomena need to be assessed in the context of several different sets of more general principles and traditions.

We begin by describing our problem very precisely, indicating why getting the correct answers is problematic. The score of the soprano melody of a short (one-minute) song, called Wine, composed by our third author (GH) and scored for soprano, WX5 Wind Controller and Digital Harmonium, is quoted (Figure 1). The score is a graphic representation of what the soprano is singing: the kind of representation which to musicians is normative. Indeed, musicians can get a reasonable idea of what the music will sound like just by looking at this part. Brahms once suggested that he could get a better sense of a piece of music from reading the score and imagining the sound of the music "in his head" than he would get from attending a concert and hearing the performance, but musicians posses this skill to varying degrees, and Brahms must surely have been exaggerating, and perhaps

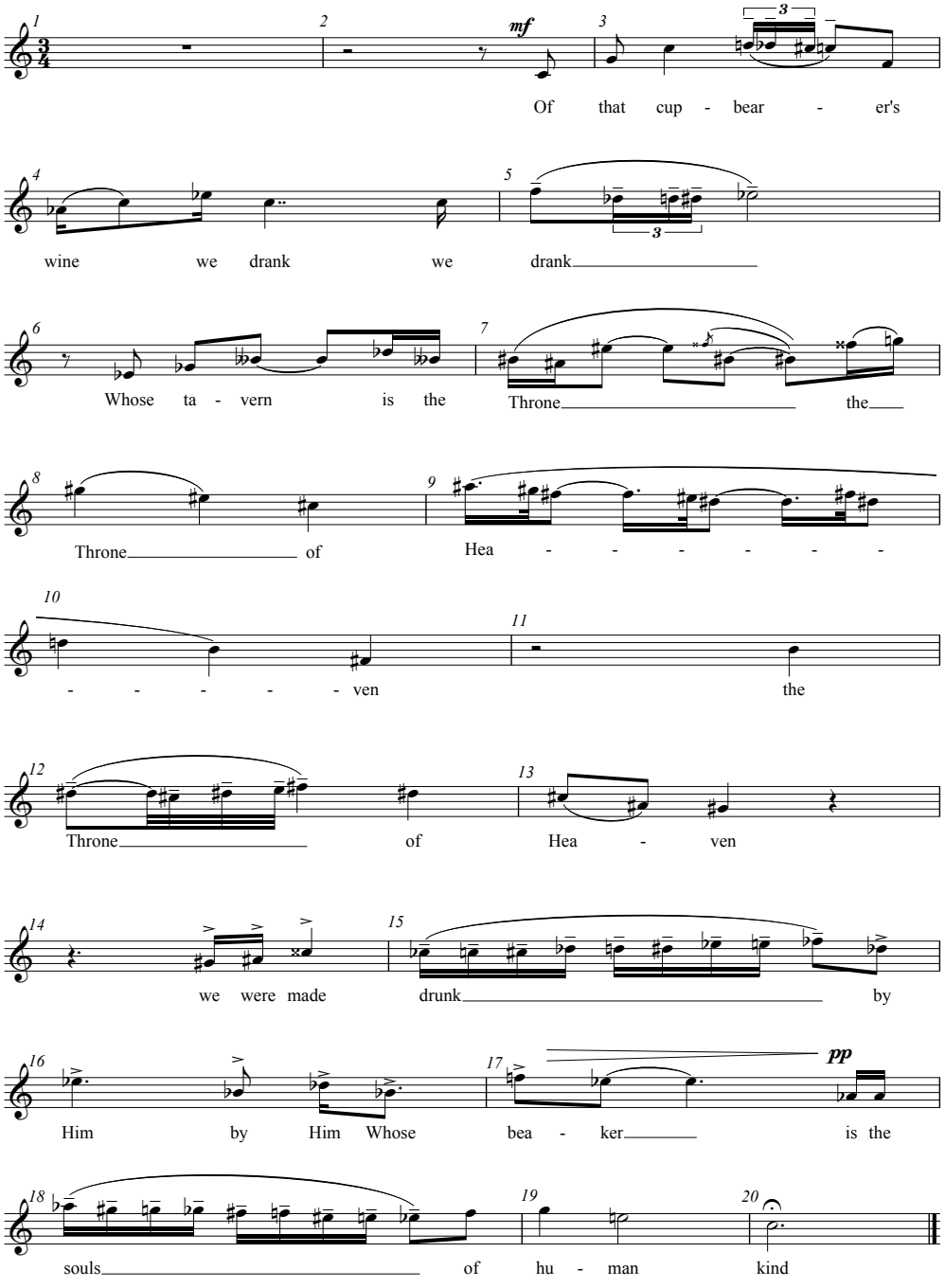

Figure 1: Soprano part from Wine by Graham Hair

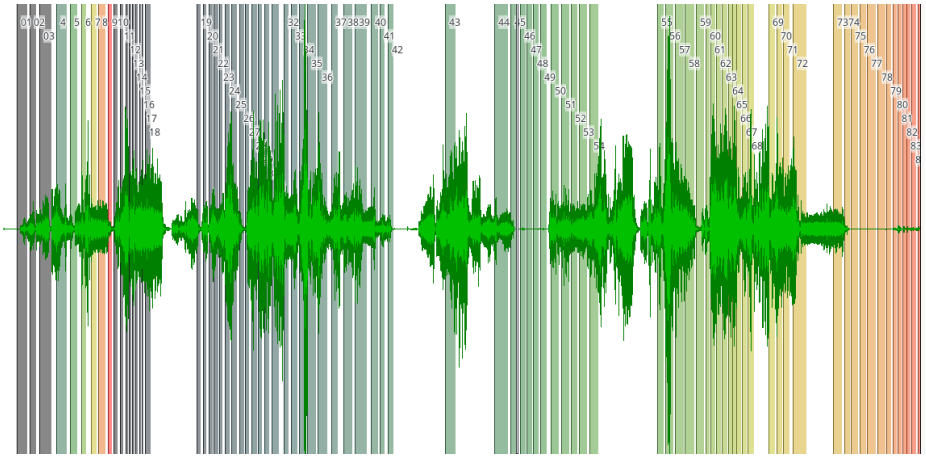

Figure 2: A wave representation of the short song Wine, each note onset and offset having been marked-up manually by the composer.

anticipating as well that the performance might not be very good. Nevertheless, the idea of what the music will sound like which musicians gain from the score is infinitely better than they can get from the next excerpt, which a graphic representation of the sound wave of a performance of Wine. This form of graphic representation is undoubtedly useful for certain purposes, but is virtually completely useless in giving musicians any precise idea of what the music will sound like.

But contemplating the score and the graphic representation of the audio file side by side may raise some questions. For example: musicologists who study performance might want to ask whether the performance is faithful to the score. That does not mean that the singer is expected to interpret the score literally or mechanically: for expressive purposes, singers diverge from a literal or mechanical interpretation virtually every time they open their mouths. But, on the other hand, that does not mean that "anything goes". What we want to know is not just that singers diverge from the literal and mecahnical, but when, why, how and more. So (to take a minute extract): with regard to the 9 pitches from the A♭5 on the down-beat of bar 18 to the E♭5 on the third beat of that bar, we may want to begin by asking whether the singer actually sings those frequencies which a literal and mechanical translation of pitch into frequency would represent. Pitch is not the same thing as frequency, of course, so we certainly need to be aware of how they differ. A first step (albeit only a first step) along that path will be to ask where, by how much, and why she diverges from a literal, mechanical ("pitch into frequency") interpretation. An experienced musician may well be able to get a sense of the answers to those questions by listening to the performance and comparing it with Figure 1, but Figure 2 may well prove to be useful for that purpose too, as technology can provide a machine-measurement of what frequences the singer actually sings from representations such as that in Figure 2, and those frequencies may then be compared with the frequencies indicated by the score.

In order to make such a machine-assisted frequency analysis possible, the musicologist needs to start by determining exactly when the singer begins and ends each note. The vertical lines in Figure 2 indicate the initiation-points for each of the 87 notes in the melody. The audio file was subjectively marked up with these initiation-points by the composer. Nevertheless, the process of

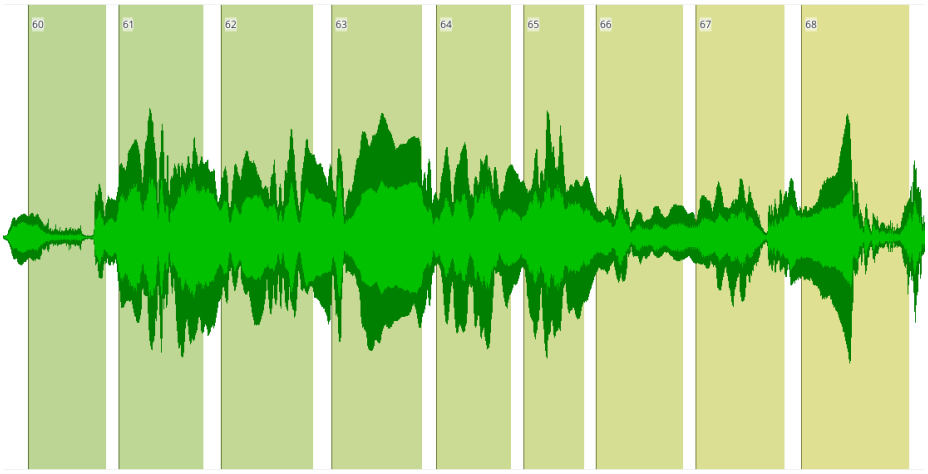

Figure 3: Detail of the file shown in Figure 2 showing only bar 18

subjective mark-up is a very rough-and-ready one. The graphic representation of the audio file requires many numbers (approximately 44100 x 60 = 2,646,000), and the task of choosing 87 of them as the initiation-points of each of the 87 notes is extremely error-prone. The initiation-points were determined by no more sophisticated a process than simply by having the cursor run through the file, and pressing "stop" at each point judged by the composer to be the initiation-point of one of the 87 notes. But to get even an approximate idea of all 87 initation-points from real-time plays-through requires several or many separate attempts accurately to place each of the 87, so as well as being error-prone the process is time-consuming and so — we may conclude — often time-wasting as well.

There is also the difficulty of even defining what is meant by the onset of a note. The audio signal of a note comprises a short transient section followed by a steady state portion. As the transient is essentially noise, pitch will not be perceived until the steady state. (In fact, transient noise is the basis of some automatic onset detectors — sudden bursts of energy across the spectrum are taken as indications that a new note has begun.) This problem is magnified in vocal music, as speech sounds can be produced with the vocal cords vibrating (voiced) or not vibrating (unvoiced), resulting in that sound's being pitched or unpitched (Reetz and Jongman 2009). As the first portion of the sound associated with a sung note may be unvoiced — or, for an instrumental note, will be noise — one could readily ask if the note has really begun when it has, as yet, no pitch.

In the field of Music Information Retrieval, the onset of the note is usually defined as the beginning of the sound(s) associated with the note, ie the beginning of the transient (Bello et al. 2005), so that the marked onsets align with the score, but there is an argument that for vocal music the onset of the note should be taken as the onset of the vowel sound (Sundberg 1994).

To illustrate the inherent difficulities of manual note segementation, Figures 3 and 4 show manually marked onsets in bar 18 of two different recordings of Wine, the first marked using the method described above (simply pressing "stop" when a new note is heard) and the second marked by slowing down the playback and employing backtracking to find more precise onset times. To be sure, the segmentation in Figure 3 was made with a particular purpose in mind: the first step in coming

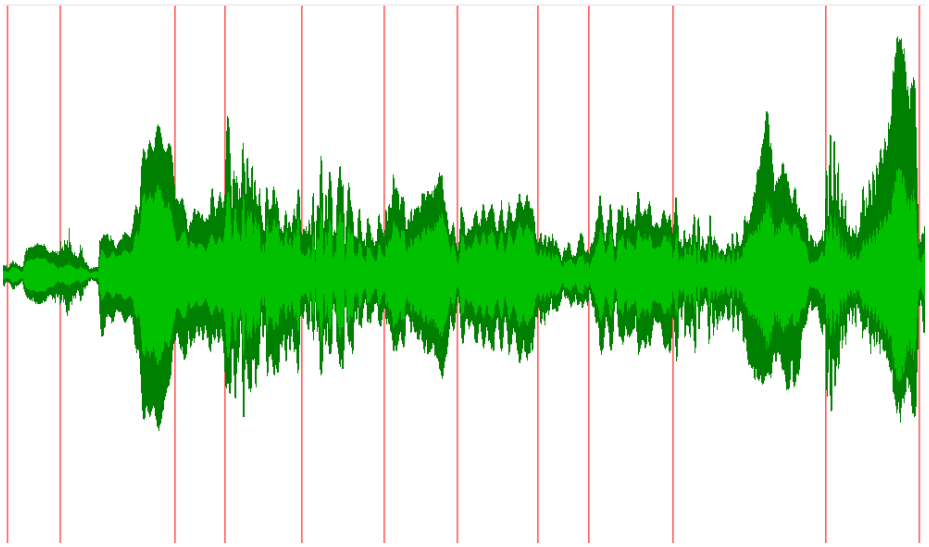

Figure 4: Bar 18 of a different recording of Wine, with onsets marked by our first author.

up with a single predominating perceived frequency value for each note of the singer's performance. It was hypothesised that this frequency value was probably best defined as an average of values somewhere across the middle of the duration of the note, so a cautious approach was adopted to the placement of note onset and termination, which are undoubtedly a little late and a little early respectively. Of course, when we have had more extensive practical experience of machine-based segmentation, we shall still have to consider different hypotheses as to where within the duration of any given note the predominant perceived frequency value (if there is one) is best located, and how to calculate it. Nevertheless, we need not throw up our hands and declare that all is continual flux and part of the eternal ineffability of life, the universe and everything, which would be counter to the well-researched phenomenon of categorical perception, one of the principles which enables us to perceive a frequency as a note in the first place.

When manually marking onsets in audio, there is the additional problem that the audio will probably contain reverberation, so the sound of the beginning of a note may overlap with that of the end of the previous one, making it even more difficult to identify the precise onset time. "Zooming in" on a short segment of the audio file, as in Figures 3 and 4 (representing bar 18 of the audio file), may seem to provide the possibility of greater accuracy, but in actual fact, what is gained in better representation of detail is lost in breadth of context, which is very important for human listeners to determine precisely where the onsets are in the audio file. The graphic representation of bar 18 of the audio file is in fact hardly more helpful than that of the whole song. Consequently musicologists have been attempting for some time to find a machine-assisted method of onset detection which is more accurate in handling the representations of the music, as provided by Figures 2 and 3. Unfortunately, "state-of-the-art" onset-detection hitherto substantially under-performs subjective detection by human listeners.

Such is the context for the reflections we offer below on the uses, abuses, extent and limits of the empirical approach to musicology.

2: Contextualising the current empirical study

Richard Parncutt's contribution to this issue of SMR provides another comment on the interaction between empirical work and established ways of thinking amongst musicians. In the course of a considerable number of publications, Prof Parncutt has produced compelling evidence that, in the light of modern work in the psychology of perception, the best, most useful way to consider the ontology ("what things are") of musical intervals is as approximate distances on a one-dimensional scale.

However, there is a long tradition amongst musicians, often said to date back to the ancient Greek polymath Pythagoras — a tradition often thought to have been confirmed in the nineteenth century by the German physicist Hermann Ludwig Ferdinand von Helmholtz (1821–1894) — that intervals are ratios, eg ratios between string-lengths (as Pythagoras proposed) or between frequencies (as Helmholtz is often considered to have demonstrated). However, Prof Parncutt has marshalled compelling psychological evidence for his alternative view. It will nevertheless be obvious that any interval which can be expressed as an approximate distance on a one-dimensional scale, will also approximate to one or other ratio, and the situation is further complicated by the fact that musicians have other ways of measuring intervals as well: notably in terms of consonance and dissonance.

So in order to disentangle these different concepts of interval, we need to consider the epistemology ("how we know what we know") of musical intervals, as well as their ontology. The reasons why musicians cling to definitions based on ratios are based on traditions in sociology, philosophy, music-theory, history, politics and maybe anthropology ... with little consideration for recent scientific studies in the psychology of perception.

3: Preamble: Plain Tales from the Coal Face, *or*: an account of some real-life experiences.

What follows are a few reflections by the authors arising from personal experience in prosecuting projects in Empirical Musicology such as the present one during the past decade, and in particular from issues which parallel the situation implicit in Prof Parncutt's work on the ontology and epistemology of interval:

1. Although the significance of evidence from Prof Parncutt's science-based work on perception has been pretty clear for some time, work in the humanities and in practice continues to proceed largely without reference to it: probably because of the tendency of scholars in humanities and of practitioners to assume that science (the "Other Culture", cf Snow 1959) is of little relevance to what they are doing;

2. More broadly, empirical findings have always to be interpreted by human beings within cultural traditions with which their own sense of identity may be tightly wrapped up: traditions bundled together with weltanshauungen which often create cognitive dissonance (in the sense of capacity to hold two contradictory opinions simultaneously) between culture and empirical observation and set limits upon the capacity of empirical observations to influence the general outlook of members of those cultural communities. We are familiar in everyday life with concepts such as "identity politics", but empirical science often comes up against "identity sociology", "identity history", "identity psychology" etc as well;

3. More broadly still: we may note that Eagleton (2015) has observed the disturbing fact that

> Culture, in the Anthropological sense — which wraps up the language, symbol, custom, belief, kinship, ethnicity etc of a whole community in a multifactorial entity (in contradistinction to High Culture embodying the values, beliefs, practices and aesthetic artefacts of a cultivated minority) — has in recent decades become, in some extreme cases, something for which individuals or social groups in some places are prepared to kill or die. Of course, Eagleton may perhaps have in mind suppressed and "radicalised" third world communities and disenchanted Western underclasses in this era of global metropolitan élites, political polarisation, disregarded and resentful electorates, deprived social minorities, radicalisation, terrorism, hate crime and the War on Terror, but the conflict between culture and empirical observation also has traction in less drastic and controversial circumstances, where life-and-death issues may not be at stake.

Musicology is just one instance of an intellectual realm in which such a clash between culture and empiricism may occur, but where questions of truth and falsehood may nevertheless arise, even if they are seldom seen in that light amongst practitioners. This may be because the field of Empirical Musicology is, in the sense of an academic sub-discipline formally thus entitled, something of a late-comer on the scene. For example, the journal Empirical Musicology Review, founded by David Huron and David Butler, began publishing only in 2006. The journal sets out to reflect the distinctive features of work in the sub-discipline, by focussing on "systematic methods, such as hypothesis-testing, modeling, and controlled observation. Theoretical and speculative articles are welcome provided they contribute to the forming of empirically testable hypotheses, models or theories, or they provide critiques of methodology." (see http://emusicology.org/about/editorialPolicies#focusAndScope).

A particular point is made on the Empirical Musicology Review website that "Authors, reviewers, and editors are required to disclose conflicts of interest at the earliest possible opportunity – for example, when a manuscript is submitted or a review assignment is accepted. Conflict of interest is defined as any competing personal, professional, or financial interest that may introduce bias into the publishing process of the journal."

Readers of such comments will no doubt sense that there's a history of some controversy, even aggravation, behind them, which the founders of the journal were seeking to head off, and in the paragraphs which follow, the present authors will sometimes refer to this history. In fact, we tend to the view that, for a long time, self-interest has been a driving force behind many aspects of the practice and the study of music, especially music criticism. The reason for this is no doubt the situation which philosopher Roger Scruton identifies:

"…the best guard against error is the freedom to question, and this freedom is at the root of academic life. That vision of "the advance of knowledge", as Bacon called it, seems to be only imperfectly endorsed in Western Universities today, where pressures to ideological conformity in the name of political correctness are constantly in the news. On the other hand, those pressures are most strongly felt in departments of the humanities, and it is sometimes argued that these departments are by their very nature, devoted less to the "advancement of knowledge" than to the propogation of moral and intellectual values. Hence it is difficult to know exactly what would be meant by teaching in the humanities in which ideological conclusions are avoided" (Scruton 2007::3).

Indeed, the implication of this observation (and indeed of the title we have given to this paper) is that there is often a mindset underpinning ideological positions which is prone to prejudice, bias and vested interest, even if there are no obvious competing "personal, professional or financial" situations involved. It's not only in The Academy that such issues crop up. Everyday life is replete

with examples. The following recent exchange on the BBC World Service provides an example. World Service Interviewer to Californian climate change denier: "What would it take to convince you to change your mind on climate change". Reply from Californian climate change denier: "I don't think anything you say could change my mind on climate change". Or the following comment from a spokesman for the US National Rifle Association after the primary-school shootings in Connecticut in 2013: "What we need to deal with a bad guy with a gun is not to take guns away from their owners. What we need is a good guy with a gun" (ie the problem is not easy access to guns, but the "darkness in the soul" — or some similar formulation — on the part of the perpetrator, and the solution to the problem of too many guns is more guns). Both speakers were part of communities with pronounced "world-views", with the result that, in both cases, the way the speakers assessed the (very specific) empirical evidence was easily trumped by the (much broader) general framework within which they looked out at the world, despite the "cognitive dissonance" (between empirical observation and prior cultural attitudes) which such attitudes require.

This mindset has also manifested itself in the development of sub-fields of musicology, and the various learned journals devoted to them, long before the advent of the Empirical Musicology Review, and which are really not new.

For example: the establishment of the discipline of historical musicology is often dated to nineteenth-century German scholarship, with such works as Carl von Winterfeld's Johann Gabrieli und seine Zeit (1834), and its "pre-history" in the post-Enlightenment musical histories in the eighteenth century of Martini and Gerbert, leading to its modern academic form and its founding bodies (such as the International Musicological Society in 1927 and the American Musicological Society in 1934). But the way in which the field of historical musicology unfolded over the course of the first half of the twentieth century, and the way in which it was configured in the minds of its practitioners, particularly its emphasis on the Western tradition and its styles and idioms, including its particular concentration on written historical documents, was one reason, amongst others, for a degree of dissatisfaction amongst those who broke away in the second half of the century to establish such bodies as the International Folk Music Council in 1947 and the Society for Ethnomusicology in 1955. Indeed, the twentieth century subsequently saw several other break-away movements, leading to the founding of, amongst others, the International Society for Jazz Research in 1969, the Society for Music Theory in 1979, the International Society for the Study of Popular Music in 1981, the Conference on Interdisciplinary Musicology in 2004, and the Journal of Music Performance Research in 2006.

But in some cases, one of the consequences has been an uneasy presentiment amongst the Establishment Status Quo that the ways of thinking espoused by the secessionists, which seem "beyond the pale" or "outside the box" from the viewpoint of hegemonic established practice, are leading to research paradigms which may not just generate a corpus of complementary studies, but place hegemony in the hands of the secessionists, consigning cherished long-accepted paradigms to the dustbin of history.

Anyone familiar with this narrative will immediately recognise common strategies deployed in "culture wars" of this kind.

Firstly, we have noticed that projects giving evidence of attitudes "outside the box" (eg utilising empirical machine measurement) sometimes meet the response (from at least some quarters) that approaches from a new direction with (by established criteria) unconventional methods cannot pretend to challenge the purview of mature fields of scholarship with recognised standards of knowledge,

insight and sophistication such as the approaches most widely-adopted in the currently hegemonic sub-fields of the discipline. But just as often, this attitude is simply a symptom that academicians in long-established disciplines feel threatened by an approach to "their" topic with a completely different methodology.

Then, when evidence of interest from Society at Large in the new approach begins to mount, we have noted that the vested-interest response is sometimes an assertion that the project covers territory which has already been investigated and understood (implication: "of no further interest to anyone with proven knowledge, insight and sophistication").

Sometimes such attitudes also hang on self-interested definitions. For example, the term "empirical" or "rational", in a general sense, is defined merely in contradistincion to "theoretical" or "logical", and may simply mean "based on, concerned with, or verifiable by observation or experience rather than theory, pure logic or rationality". To consider a specific instance: folk-song transcriptions in the early twentieth century by Percy Grainger and Bela Bartok attempted to transcribe folk-song performances by singers from various aural traditions, using the conventional notation of the "common-practice" tradition. These are empirical in the aforementioned general sense: that is, Grainger and Bartok tried to capture the results of objective observation by writing down more precisely than earlier folklorists had done exactly what their subjects actually sang to them, which might mean a written score full of more detailed expressive melodic nuances and more irregular metrical and durational rhythmic patterns than in the published editions of earlier transcribers, who may have written down more regular patterns which they imagined as some kind of underlying regular framework over which such nuances were laid, perhaps by analogy with rubato-infused performances of more regularly-notated music in the concert tradition. From a Graingerian or Bartokian perspective, Cecil Sharp's transcriptions of English folk songs, for example, are sometimes described as examples of romanticisation, bowdlerisation or regularisation of his material.

Nevertheless, in the twenty-first century, investigative methods are often considered empirical principally (or only) if they use current scientific instrumentation to make the kind of machine-driven objective measurement which Grainger and Bartok made "by ear" and "by hand" (so to speak). Again, we have sometimes encountered an attitude amongst academicians regarding themselves as having proprietorship over (in this case) "empirical" methodology, feeling threatened, and treating with suspicion, practitioners who approach "their" topic with a different technology-based methodology which was not available in Grainger's or Bartok's day. But in fact, contemporary practitioners are simply continuing an empirical tradition with twenty-first-century methodologies and technologies. No-one thinks that Einstein's work should be regarded with suspicion because Newton pre-empted him, and that Newton's creation of modern physics in the seventeenth and eighteenth centuries somehow means "That's already been done!". Einstein's work is simply a twentieth-century development of an earlier-established tradition along new lines.

And when, finally, thinking "outside the box" achieves something recognised by Society at Large as a significant contribution to knowledge, the vested-interest response is sometimes to proclaim it inconsequential. For example: music has always, since time immemorial, attracted the interest of scientists, and in some earlier times, was often — or even usually — regarded as a science. But in the contemporary Academy, music is classed as one of the humanities, despite the continuing interest from scientists. Interest from scientists does not however mean that we should now cease to regard music as one of the humanities, but simply that we should not defer to an attitude of proprietorship amongst academicians whose lifespan has fallen within the period during which music has always

been considered as one of the Humanities, especially when such proprietorship takes the form of the aforementioned "inconsequential" proclamation, such as by comparing the efforts of twenty-first-century empirical musicology to counting the number of F sharps in the corpus of Mozart String Quartets or the number of definite articles in the complete works of Shakespeare, ie characterising the research involved as a trivial matter appropriate only to shop-keepers, accountants, "geeks" and "bean-counters" (in contemporary academic parlance).

Often the "inconsequential" label derives from some form of caricature: that is, misrepresentation in which, nevertheless, there is a small grain of truth. An example would be the assertion, in response to studies with a psychological basis, that empirical musicology investigates perception but ignores culture. No music psychologist who has conducted studies in musical perception has ever thought culture irrelevant, but all the same, caricature is dragged out to suggest that such a truism has somehow been bypassed.

Again: consider the case of the assertion that "every human being is a unique individual, and everyone hears differently": not a critical insight but a trusim which no empirical musicologist has ever doubted. Every human being is indeed a unique individual, but nevertheless, the achievement in the course of Western cultural history of the so-called "common practice tradition" has been made possible only by a degree of fraternity (of perception and culture) between such unique individuals. And if Dmitri Tymozcko (Tymoczko 2011) is right, what he calls the "Extended Common Practice Tradition" is based on common ways of perceiving and understanding music which can be traced back over a millennium, and which also have many things in common with traditions of popular music and many forms of extra-European musical practice.

Moral philosopher Harry Frankfurt, Professor Emeritus of Princeton University, analysed the prejudice, bias and misrepresentation embodied in such deliberately misleading rhetoric in a celebrated recent essay which appeared for twenty-seven weeks in the New York Times Best Seller List in 2007: entitled On Bullshit (Frankfurt 2005). It perhaps says something cheerless about the nature of the twenty-first century academy, that such a title, clearly designed in ironic parallel with those of such ancient philosophical works as Lucretius's On the Nature of Things, should be devoted to such a topic. But the concept has a venerable history; Treitler (2011:34–35) cites an example in the music-theoretical works of Aurelius of Réôme in the ninth century, and it plays a considerable role in the novels of Charles Dickens in the nineteenth century, although Dickens has a different term for it, namely "humbug". Indeed, Frankfurt's was foreshadowed 25 years earlier: Max Black's The Prevalence of Humbug and Other Essays (Black 1985). Frankfurt's use of the term "bullshit" is not that of everyday colloquial usage, as a generic term of abuse, but as a quasi-technical term, meaning approximately "the misrepresentation of vested self-interest in a way which endeavours to make it appear as objective and universally-valid truth". As Wikipedia puts it: the underlying social situation is that the bullshitter's sole concern is personal advancement and advantage to his own agenda.

To take a local example: in a paper presented to a recent Musica Scotica conference, Aberdeen-based musicologist Ronnie Gibson noted that, in the course of his study of the nineteenth-century Scottish fiddle music repertoire, he had often encountered the term "decline" to characterise the period. He noted that the nineteenth century was nevertheless a period of considerable creative activity in the field, and suggested that the characterisation of the period in that way was primarily an historiographical construction invented by twentieth-century writers intent on seeing themselves as revivers of a noble eighteenth century tradition assocociated with the name of Neil Gow and

others. Indeed the prosecution of agendas of this kind is virtually a constant of the historiography of music, driven by the vested interests of the writers, not by an assessment of the objective truth about the material in question, ie "bullshit" in Frankfurt's precise definition.

Given the appearance of such corrupt rhetoric even in the groves of academe, it is not surprising that under the pressure of the forces in The Academy, it has begun to corrupt even such sacred cows as peer-review. For instance, David Horton, the eminent medical researcher and editor of The Lancet famously had this to say on the subject (even though he is talking here about the supposedly more objective discipline of science, not about humanities or artistic practice):

"The mistake, of course, is to have thought that peer review was any more than a crude means of discovering the acceptability – not the validity – of a new finding. Editors and scientists alike insist on the pivotal importance of peer review. We portray peer review to the public as a quasi-sacred process that helps to make science our most objective truth teller. But we know that the system of peer review is biased, unjust, unaccountable, incomplete, easily fixed, often insulting, usually ignorant, occasionally foolish, and frequently wrong." (Horton 2000).

These are all essentially examples of attitudes with which everyone is familiar in everyday life, and which have now often come collectively to be characterised as "Bubble Culture", in which the vested interests of small social groups gain ascendancy over empirical truth. The British population is very familiar with the "Westminster Bubble" (the isolation of the attitudes of politicians in Westminster from those of members of Society at Large) but it is a widespread phenomenon in many fields, as the global furore created by Frankfurt's celebrated essay suggests. Bubble Culture in music often seems driven by what we have designated above as the aspirational ideology embodied in some aspects of High Culture in the form of the Nirvana Fallacy: which may be defined as the putting of unrealistic, idealized alternatives in the place of achievable real things, or being driven by the assumption that there is a perfect solution to a particular problem. Its essence is a kind of utopianism along the lines of: "Don't let How Things Are stand in the way of How Things Ought To Be", or even, in an extreme versions of that thought, in the colloquial "journalese" expression: "Don't let inconvenient facts stand in the way of a good story".

4: The Uses of Empiricism

Some of the approaches outlined in our introductory comments have clearly been influenced by readings in inter-disciplinary sources: History, Sociology, Politics, Literature, Languages, Philosophy. The following sections return to science in various forms and the issues raised in our foregoing introductory sections. Although the production and performance of music is not itself configurable as a form of science or empirical research, that does not mean that music cannot take advantage of research in History, Sociology, Politics, Literature, Languages and Philosophy. Realising some degree of fraternity across disciplines seems virtually to require it.

Likewise, overcoming any temptation to fall into the trap offered by the Nirvana Fallacy may be facilitated by conjoining insights derived from practice to research in both science and the humanities: in the former case to take a purely empirical approach to the acquisition of data for analysis, and to achieve this, accurate, automatic frequency and onset detection are a sine qua non. Discussion of the aesthetic issues might then at least proceed based on verifiable data rather than opinion, speculation and vested self-interest.

Empirical methods familiar in the sciences have been applied to music analysis in a number of fields besides musicology — for example neuroscience, psychology and sociology (Clarke and Cook

2004; Huron 2006). Research of this nature, both in such interdisciplinary fields and empirical musicology, attempts to tell us not only what is happening in the music, but how and why.

In the following sections, we describe the empirical methods that have been deployed in gathering data for performance analysis of the microtonal (19-EDO) music here under consideration. It is unsurprising that there is little ready-made software available which is sufficiently flexible to accommodate a "non-standard" temperament. As well as performance analysis, we have discovered that there are other issues to be addressed in developing and using software supporting the production of music such as this. Consequently we direct our efforts to plugging the gap in all of the aspects which benefit from modern technological assistance, not simply confining the effort to sound production and measurement.

There is a second reason why engineering and scientific research needs to be brought to bear on the problem of empirical performance measurement. As we shall show, the existing state-of-the-art simply isn't as good as the human ear when it comes to extracting musical (as opposed to audio) information from the sound of a performance. We should not be surprised that this is so. It is presumption to suggest that in a very few decades there could be developed a system as acute in musical terms, which is a human factor, as that achieved by many millennia of human evolution. But even those who should know far better are often seduced by the idea that modern computers applied to processing digital recordings can extract musical information as easily as a policeman's radar measuring the speed of a car. We can track the pits of a blu-ray disk flying past a laser read head at the rate of tens of millions per second, or hurl a probe so that it bumps gently into an asteroid, so surely extracting musical information must be a relatively simple problem? This specious confusion of music with audio ignores the fact that the audio samples we use to represent an instance of the musical performance is related to the musical content in the same way that Morse code is to literature. It can convey one manifestation of a performance to an arbitrary degree of accuracy while in itself telling us nothing whatsoever about it.

First, then we shall examine the tools which are somewhat mechanical, and where bespoke solutions are sought to bring a portable, generic tool our particular microtonal musical question. Later we shall move on to consider the tools one step higher up the musical "food chain" and present a method of automatically segmenting a performance into notes, as has already been described, with greater reliability and accuracy than has been available heretofore.

5: Making Measurements and Made-to-measure

Applying signal processing technology in the provision of rehearsal aids for expert musicians practising microtonal pieces has been covered elsewhere in this issue, so we shall consider next the technological aids and interventions available to the composer for the preparation and dissemination of microtonal scores to the prospective performers, then move on to describe the methods under development to enable an objective analysis the performance outcome.

One might assume, again speciously, that the preparation of a performing edition is simply a typesetting exercise. In fact, the rehearsal and proof-reading processes are naturally assisted a great deal by the preparation, alongside the printed parts, of an audio rendering, however devoid of expression. This is even more true when the music is to be performed in a temperament outwith the performer's everyday experience. The provision of a computer-generated proof moves the cognitive load away from a note-by-note approach to retuning and permits understanding to commence more-or-less immediately at the motivic and phrase level.

Although it should go without saying, music and audio are not the same. Common usage is shifting to disguise this distinction: even practitioners might be heard referring to "music players" (when they mean audio players), the "music industry" (recording industry) or ask where they have left their music (score, one would hope). Surely not even the most luddite musician would deny there is at least some value in using technological tools to record rehearsals or performances for the purpose of self-analysis (Bailey et al. 2008b). Similarly, it is a valuable exercise to analyse the acoustic outcome (recorded sound) of a music-making process (performance or rehearsal) in order to provide objective evidence in support of a musicological hypothesis. So long as the recording process is unobtrusive and the analytical tools accurate and appropriate, any argument against doing so can only be based on a self-interested desire to maintain long strived-for and dearly-held prejudices.

Having conned a part then, it is, to say the least, instructive to measure what the performer produces, and not just once but at a series of performances over time as familiarity with the work grows. If the composer's intent is to produce a work for performance and reception by the audience, ipso facto there is an interest in the extent and nature of deviations from a mechanical interpretation of what is written down in the notation. One avoids emotive terms such as "accuracy of tuning/rhythm" because, obviously, music resides, in considerable part, in the deviation between what is performed and notated (otherwise all of the great performances would be given by computers, and we know the opposite to be the case). Unless the preservation of an opinion regardless of the facts is the ultimate aim, it is still potentially very useful to collect accurate performance data.

In summary, we desire to:

1. compose music;
2. print out performing editions;
3. maybe "audio proof-read" the score, certainly be able to pass rehearsal aids to performers;
4. learn by performance analysis what the performers are actually doing to accelerate the development and acceptance of idiomatic (in this case) 19-EDO repertoire.

In the following sections we describe the tools used to achieve these ends. Because the corpus we are creating and examining in this context is not in 12-EDO temperament, there are far fewer proprietary applications which can be used "out of the box". In any case, such applications are almost always predicated on narrow usages and are barely extensible. For example, the primary purpose of score editing software is graphical layout, and even if it is possible to produce audio proofs with a non-standard temperament, it is vanishingly unlikely that the score can be annotated automatically with arbitrary graphics to indicate measured or intended performance practice beyond those normally found in a conventionally printed copy. So, we adopt a tool-based approach (Bailey et al. 2008a) by combining a series of small, testable and therefore (at least potentially) reliable programs to produce performance data which may then be combined with the score in a database (Pullinger et al. 2009) tailored for the storage of performance information, not simply the audio and textural instantiations of the music under consideration. This flies in the face of the "milk comes from supermarkets" attitude towards computer software adopted usually by people who have never written any, and while it is an undeniably monumental task to produce a monolithic, large scale "one-stop shop" for performance capture and analysis, a proper tool-based approach, along with the use of those tools already available for the purpose, mean that meaningful results can be produced from bespoke software systems.

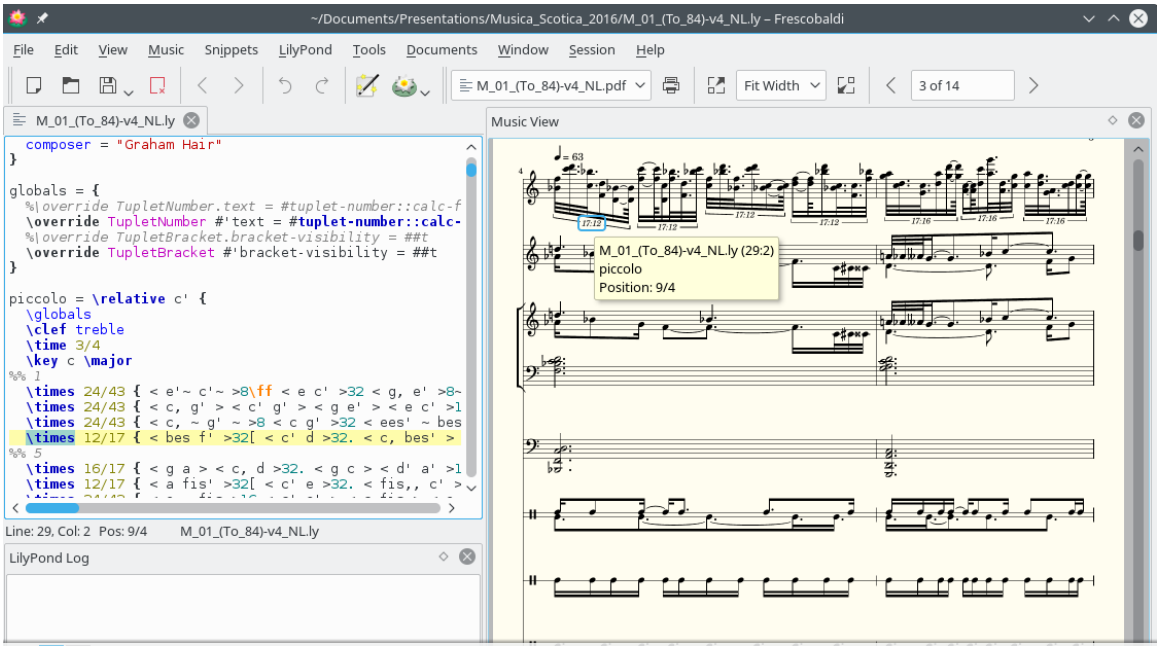

Figure 5: A Frescobaldi session editing a 19-EDO composition by Graham Hair. Note that selecting a feature in the score preview window automatically selects the corresponding Lilypond command

6: A Microtonal Toolchain

Notating the kinds of microtonal score under consideration here, 19-EDO, isn't problematic for the Lilypond music engraver. The 19-step scale is isomorphic in several respects with the 12-step one, and no new accidentals are needed.

Figure 5 shows an extract from our third author's composition Enneakaidekaphonic Study No 1 notated in Lilypond. This piece uses strict time values to notate gradual changes of "felt" tempo in one of the parts set against fixed tempi in the others, a notational practice not uncommon in twentieth-century music such as that of Elliott Carter and others. Additionally, using Graham Breed's "regular" macro (`http://x31eq.com/lilypond/`) extends Lilypond's MIDI renderer to produce, along with the note-on messages, pitch-bend control signals which cause a compliant MIDI renderer to produce output at the intended. Adding

```
tuning = #19
\include "regular.ly"
```

to the start of the Lilypond file results in the pitches being bent in such a way as to render them in 19-EDO (assuming the file regular.ly is in Lilypond's search path).

Unfortunately, this procedure doesn't work very well for polyphonic scores, which in MIDI terms means parts with voices simultaneously sounding in the same channel. Although MIDI provides a pitch-bending command, it operates on the whole channel at once, so that when two notes are sounded simultaneously followed by their appropriate (different) pitch-bend corrections, the last one "wins" and applies to both notes. There is a facility to alter the pitch of a note separately,

the "MIDI tuning standard", but this is considered an extension and is not well supported by (at least hardware) synthesisers. Breed also supplies an addmts program which can reinterpret the standard MIDI stream and replace the pitch-bend messages with device-dependent retune ones, and the combination of this and the "regular" lilypond MIDI tuning scheme script makes it possible to produce well-tuned proof-reading and rehearsal-aid renderings direct from the Lilypond notation.

We have already observed that an interesting part of musical performance lies not in its similarity to the score, but in its deviation from it. Obviously, this includes such things as the addition to the score of ornamentation (trills etc) by performers, but much else besides. To record all of the necessary data in a single resource necessitates storage of both the score and the recorded performance parameters. To this end, we use an extension to the popular relational database, PostgreSQL, called SpofF, because it is uses a "spiral of fifths" concept to record the pitch of note events. These are associated with the recorded performance attributes on a per-performance basis, and thus a corpus of performances could be queried individually or as a whole, the results being presented graphically combined with the score. The most comprehensive description of the system is given by Pullinger (2010).

Internally, the Sprial-of-Fifths permits any diatonic scale to be represented by recognising that the "circle of fifths" used in 12-EDO descriptions really doesn't "join up": one might insert B𝄫 before F♭, or F𝄪 after B♯ and continue arbitrarily before determining any notes to be genuine enharmonic equivalents. The representation connects pitch class with performance pitch using temperament. The SpofF specification has the facility to represent some tunings, such as quarter-tones, as multiple offset spirals where musical semantics call for this.

The database extensions (including the new data types necessary to support the new representation and programs which populate the database from PerfomanceML source) are made available in a package called Thomas (originally, Thomas the Note Engine after the series of books by Rev'd Wilbert Awdry). A second package, ARNE (Automated Retrieval of Note Events), written in Python, permits simple algebraic operations on metrical time and intervals within the database. For example, the following advances two beats from bar 6, beat 2 in $\frac{3}{4}$ time and finds the pitch a Major third above G4:

```
>>> import spoff
>>> addDuration ({'bar':6, 'beat':2, 'division':0},
                 {'bar':0 , 'beat':2 , 'division':0} , 3)
{'bar':7, 'beat':1, 'division':0}
>>> pitch2text ( addInterval(text2pitch('G4'), text2interval('M3')) )
'B4'
```

By maintaining a fifths-based representation it is difficult to confuse the algorithm even with the most unlikey requests. The following requests the note a doubly-dimished sixth above B♭ 4:

```
>>> import spoff
>>> pitch2text ( addInterval(text2pitch('Bb4'), text2interval('dd6')) )
'Gbbb5'
```

Combining this with the SQL queries, augmented with Spiral-of-Fifth types, arbitrary and powerful queries may be written.

These tools have proved useful in storing and manipulating musical information in this and previous projects, but their novelty comes from extending and generalising existing technologies rather than taking a substantially new approach. We now describe a completely new technique which has been adopted to (predominately) automate one of the most tedious aspects of the empirical performance process: that of automatic note segmentation.

7: An Empirical Approach

Despite the fact that computers are increasingly able to process vast amounts of data in ever shorter times, their use in music analysis is limited, as accurate data acquisition by computer has not improved at the same rate. As a result, manual data collection techniques are used, with the data then being analysed by computer. Due to the difficulty and time-intensiveness of manual techniques for collecting musical data for computer analysis, they have mostly been limited to timing applications, like beat tracking, which involves repeatedly pressing the space bar to mark every beat in a piece of music (this is actually even more tedious than it may appear — in order to do this accurately the person pressing the key has to become extremely familiar with the piece and so know when accelerandos and rallentandos will occur), or segmenting an audio file into individual notes by manually marking the beginning of each note.

Note segmentation of audio files is an essential first step in performance analysis. Manually marking-up audio files is exceptionally time consuming, so a tool for automatic segmentation would be invaluable, as it would allow a great deal more data to be analysed and larger sample size means more robust conclusions can be drawn. Such a tool could function by detecting the onset of notes, possibly in combination with pitch tracking. However, there is currently no available software which can reliably detect non-percussive onsets (Milligan and Bailey 2015).

Several onset detection algorithms were tested by Böck et al. (2012). The audio dataset was comprised of a variety of instrumentation including percussion and bowed strings. Although the results were not broken down by instrumentation, the F-measure[1] (Witten et al. 2011) for each of the algorithms tested never exceed 80%.

MIREX (Music Information Retrieval Evaluation Exchange) runs annual competitions in, amongst others things, automatic onset detection. The results are available online at `http://www.music-ir.org/mirex/wiki/MIREX_HOME`, the most recent of which are broadly similar to the experimental results obtained by Milligan and Bailey (2015): the algorithms simply do not perform well for solo voice, possibly due to features particular to vocal music. A single note could contain more than one speech sound (in which case there may appear to be many onsets) or one speech sound can last for more than one note (onsets may not be well enough defined to be detected by the program). Vibrato may also cause erroneous detections.

A robust and reliable onset detector should be able to deal with all kinds of music, including extra-European music and that which employs non-standard tuning. Of course, instruments which are not tied to any particular temperament or tuning system are often those with non-percussive onsets (eg voice, unfretted (bowed) strings), which only serves to compound the problem.

Although manual segmentation is inefficient, it is generally not difficult for a person to tell where a note begins; that automated onset detectors perform so poorly suggests they are taking the wrong approach and that conventional signal processing techniques are not the right tools for the job.

8: Shortcomings of the Fourier transform

A common method for representing a time series in the frequency domain is the Fourier transform. Whilst it is useful in many fields, it is not suitable for musical applications for several reasons. Firstly, it requires a stationary signal. In practical terms, this means the signal must have a constant spectrum for longer than it would ever be in music, especially considering that the audio of a single note might consist of a noisy transient region, followed by a steady state (which may be very short) or a deliberate continuous pitch variation (eg vibrato or glissando).

The form of the Fourier transform often used in musical analysis is also problematic. The short time Fourier transform (STFT) is often used to calculate fundamental frequencies in music (Crochiere 1980; Devaney and Ellis 2008; Flanagan and Golden 1966). This is a form of the discrete Fourier transform (to process sampled data), which uses short time windows (typically 23ms) and breaks the signal into frequency components in evenly spaced bins. However, human pitch perception is approximately logarithmic, so the Fourier transform does not adequately mimic this: there is too little information at low frequencies and too much at high frequencies.

The constant-Q transform attempts to rectify this by adapting the STFT to give its output in logarithmically-spaced frequency bins (Brown 1991) so that the relationship between centre frequency and bandwidth for each is constant (hence constant-Q). This is more appropriate for musical applications (Brown 1992; Smaragdis 2009), as it preserves information about relative pitch changes. In a spectrogram drawn using constant-Q data, any two notes an octave apart would appear the same distance apart on the graph.

Like the STFT, the constant-Q transform uses windowing. It achieves the constant-Q for each bin by changing the length of the window depending on the frequency component being analysed, with window length decreasing as the centre frequency increases.

Another alternative is the wavelet transform (Addison 2002; Nanavati and Panigrahi 2004). A multi-scale method for analysing a signal, the wavelet transform is calculated by convolving the signal with a suitable wavelet. A wavelet is a short wave; common wavelets include the Haar wavelet (one period of a square wave), Ricker wavelet (second derivative of a Gaussian) and the Morlet wavelet (a complex, Gaussian-windowed wavelet). The location and dilation parameters of the chosen wavelet can be varied so that the output can have any resolution at the desired scale (ie you can 'zoom in' on a specific part of the signal).

All of the above methods fall foul of the entropic uncertainty principle due to their use of windowing. Generally speaking, uncertainty principles express the inability of two simultaneous sharply localised representations of the same function to exist (Folland and Sitaram 1997; Ricaud and Torrésani 2013). In signal processing applications, the uncertainty principle refers to the deterioration of time resolution as frequency resolution is increased or vice versa. That is, the more precisely we measure the time at which something occurs, the less precisely we know its frequency. This is obviously something of a problem for musical applications, as it is often necessary to know both the precise onset time and pitch of a note.

In the context of the multi-scale wavelet transform, where one can choose to have a high resolution anywhere in either the time or frequency domain, the effect of the uncertainty principle is to reduce the resolution in one domain as the resolution in the other domain is increased but not both, as described above.

Similarly, for the STFT, the (fixed) window length determines the resolution at all scales. Choosing a longer window would allow greater frequency resolution at the expense of time resolution and

vice versa.

The constant-Q transform adapts the window length in a predetermined way — lower frequencies have longer windows (increasing the frequency resolution) and higher frequencies have shorter windows (so the frequency resolution is decreased).

Software which does not require windowing and where transformation between time and frequency domains is not necessary — that is, software which takes a fundamentally different approach to the problem — could side-step the restrictions of the uncertainty principle.

9: SMRG's software

The Science and Music Research Group's software for note onset detection was developed by first considering the auditory system. Pitch and time perception occur in the cochlea (Gomez and Stoop 2014), as the inner hair cells detect vibration causing, auditory neurons fire (Pickles 2012).

The organ of Corti, found within the cochlea, contains the basilar membrane on which lie the inner hair cells and the outer hair cells.

As sound is transmitted to the cochlea, a travelling wave appears across the basilar membrane. The point of greatest amplitude of the wave depends on the frequency of the sound, with low frequencies represented at the apex of the membrane and high frequencies at the base. Békésy's discovery of these travelling waves in the 1920s (Olson et al. 2012) confirmed the basilar membrane tonotopic map of frequency, which had been hypothesized in different forms by various other scientists, notably Helmholtz with his resonance theory of hearing (Finger 1994).

Although the role of the basilar membrane in performing frequency discrimination had been substantiated, the mechanism by which the membrane's passive response achieved the acuity of human pitch perception was unknown.

The idea of an active process, which adds energy to the basilar membrane response, was first mooted in the 1940s (Gold 1948) but was not confirmed until the 1970s, with the discovery of otoacoustic emissions (Kemp 1978, 1979): the outer hair cells, which lie on the basilar membrane, oscillate both in response to sound and spontaneously, resulting in the emission of sound from the cochlea.

The precise mechanical mechanism by which this works is a more recent discovery (Eguíluz et al. 2000; Hudspeth 2008). The Hopf bifurcation is a mathematical concept, characterising a system which goes from stability to instability as some parameter crosses a critical value and the eigenvalues become purely imaginary (Kuznetsov 2004). In the auditory system, behaviour of the outer hair cells can be described by the Hopf bifurcation. When they are operating on the stable side of the bifurcation, the response of the basilar membrane is amplified, tuned and compressed; on the unstable side they oscillate spontaneously and otoacoustic emissions occur.

Our software for note onset detection, currently in development, takes as its basis a form of the Hopf equation, which is driven by the sound input. An array of these detectors, each tuned to a characteristic frequency, can be used as onset detectors and pitch trackers. The problems which arise due to entropic uncertainty when using the Fourier, wavelet and constant Q transforms (as discussed above) do not apply here, as the times and frequencies are not being measured: they are directly accessible state variables of the driven system.

Some preliminary results are shown in Figure 6. This shows the output of thirteen detectors, tuned to the frequencies corresponding to F4 to F5 in standard 12EDO tuning. The sound input is a chromatic scale from F4 to F5 played on a digital piano. It can be seen that each detector reacts

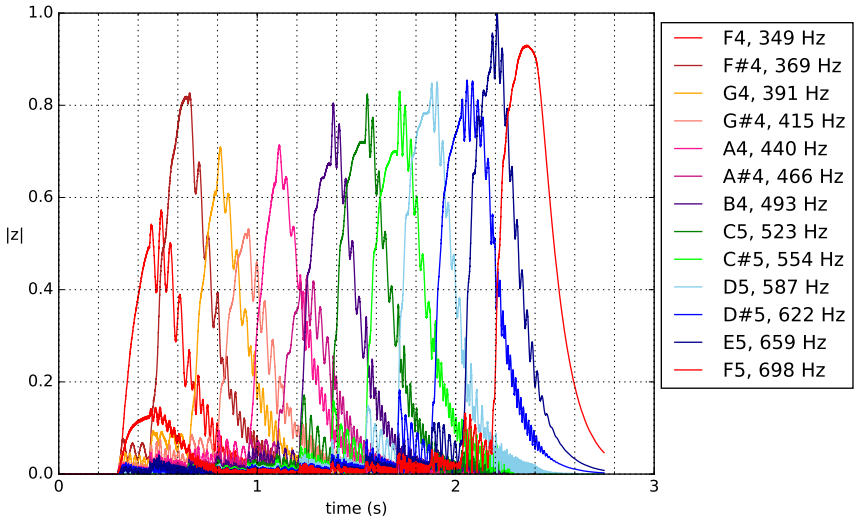

Figure 6: Thirteen detectors, tuned to a chromatic scale, responding to a chromatic scale played on a digital piano.

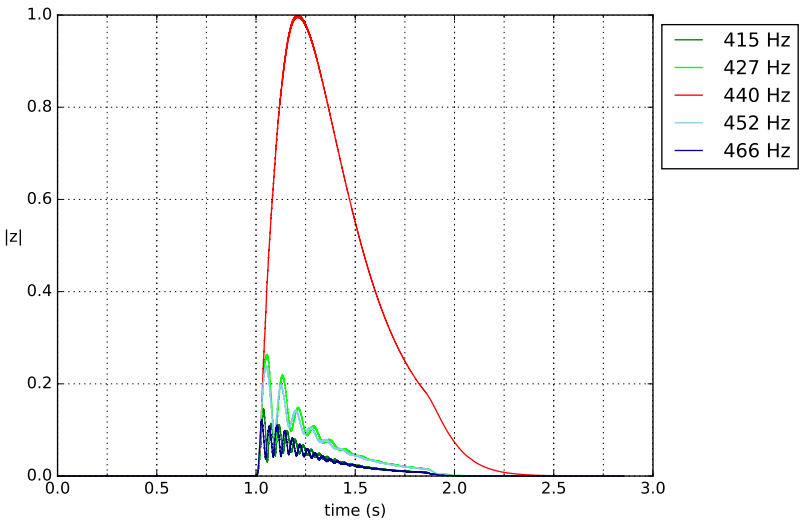

Figure 7: Output of five detectors, with characterictic frequencies corresponding to A♭4 to A♯4, increasing by quartertones. The sound input is an A4 (440 Hz) played on a digital piano.

strongly to sound at its characteristic frequency and rejects neighbouring semitones. Harmonics, such as they can be seen here, do not cause problems either — at the onset of the F4, it can be seen that the F5 detector reacts slightly as well, but this would be below the threshold for onset detection.

Figure 7 shows the output of five detectors, with characteristic frequencies corresponding to A♭4 to A♯4, increasing by quartertones. The sound input is an A4, again played on a digital piano. The sharp frequency selectivity is even more pronounced here. To build a filter with such strong rejection of propinquitous frequencies with conventional digital filtering techniques would require a roll-off of 200-300 dB per octave.

This figure also shows the quick response time. The sound starts at exactly one second and the detectors react virtually instantly. Preliminary results of using template matching to find onset times in the data shown in Figure 6 has errors no greater than 14ms. Two sounds that occur within 30ms of each other will be perceived as simultaneous (Moore 2012) and these results are well within that margin. For comparison, a Fourier transform with a 14ms window could only resolve adjacent frequencies separated by at least 71 Hz.

This software will be useful for research in performance practice, with no limitations of non-standard tuning and could also be used as a training tool for musicians.

Notes

[1] The F-measure is a statistical measure, which takes into account both the precision (the proportion of detected onsets which were correct) and recall (the proportion of correct onsets which were found).

Bibliography

Addison, Paul S. (2002). The illustrated wavelet transform handbook. (New York/London: Taylor & Francis). [Cited on page 17].

Bailey, Nicholas, Douglas McGilvray, and Graham Hair (2008a). Musically significant, automatic localisation of note boundaries for the performance analysis of vocal music. (Proc Conf Interdisciplinary Musicology, Thessaloniki). [Cited on page 13].

Bailey, Nicholas, Douglas McGilvray, Graham Hair, Ingrid Pearson, Amanda Morrison, and Richard Parncutt (2008b). The rosegarden codicil: Rehearsing music in nineteen-tone equal temperament. (Proc International Computer Music Conference). [Cited on page 13].

Bello, Juan Pablo, Laurent Daudet, Samer Abdallah, Chris Duxbury, Mike Davies, and Mark B Sandler (2005). "A tutorial on onset detection in music signals", Speech and Audio Processing, IEEE Transactions on 13(5):1035–1047. [Cited on page 4].

Black, Max (1985). The prevalence of humbug and other essays. (Ithaca, NY: Cornell University Press). [Cited on page 10].

Böck, Sebastian, Florian Krebs, and Markus Schedl (2012). "Evaluating the online capabilities of onset detection methods", in Ismir. 49–54. [Cited on page 16].

Brown, Judith C. (1991). "Calculation of a constant-Q spectral transform", The Journal of the Acoustical Society of America 89:425–434. [Cited on page 17].

——— (1992). "Musical fundamental frequency tracking using a pattern recognition method", The Journal of the Acoustical Society of America 92:1394–1402. [Cited on page 17].

Clarke, Eric, and Nicholas Cook, eds. (2004). Empirical musicology: Aims, methods, prospects. (New York/Oxford: OUP). [Cited on page 11].

Crochiere, Ronald E. (1980). "A weighted overlap-add method of short-time fourier analysis/synthesis", Acoustics, Speech and Signal Processing, IEEE Transactions on 28:99–102. [Cited on page 17].

Devaney, Johanna, and Daniel P. Ellis (2008). "An empirical approach to studying intonation tendencies in polyphonic vocal performances", Journal of Interdisciplinary Music Studies 2(1-2): 141–156. [Cited on page 17].

Eagleton, Terry (2015). "The death of god and the war on terror", Oxford Left Review 14(2). [Cited on page 6].

Eguíluz, Víctor M., Mark Ospeck, Y. Choe, A. J. Hudspeth, and Marcelo O. Magnasco (2000). "Essential nonlinearities in hearing", Physical Review Letters 84(22):5232. [Cited on page 18].

Finger, Stanley (1994). Origins of neuroscience: A history of explorations into brain function. (Oxford/New York: Oxford University Press). [Cited on page 18].

Flanagan, James L., and R. M. Golden (1966). "Phase vocoder", Bell System Technical Journal 45(9):1493–1509. [Cited on page 17].

Folland, Gerald B., and Alladi Sitaram (1997). "The uncertainty principle: a mathematical survey", Journal of Fourier analysis and applications 3(3):207–238. [Cited on page 17].

Frankfurt, Harry (2005). On bullshit. (Princeton, NJ: Princeton University Press). [Cited on page 10].

Gold, Thomas (1948). "Hearing. II. The physical basis of the action of the cochlea", Proceedings of the Royal Society of London B: Biological Sciences 135(881):492–498. [Cited on page 18].

Gomez, Florian, and Ruedi Stoop (2014). "Mammalian pitch sensation shaped by the cochlear fluid", Nature Physics, 530–536. [Cited on page 18].

Horton, Richard (2000). "Genetically modified food: consternation, confusion, and crack-up.", The Medical Journal of Australia 172(4):148. [Cited on page 11].

Hudspeth, A. J. (2008). "Making an effort to listen: Mechanical amplification in the ear", Neuron 59(4):530–545. [Cited on page 18].

Huron, David (2006). Sweet anticipation: Music and the psychology of expectation. (Cambridge, Mass/London: MIT press). [Cited on page 12].

Kemp, David T. (1978). "Stimulated acoustic emissions from within the human auditory system", The Journal of the Acoustical Society of America 64(5):1386–1391. [Cited on page 18].

——— (1979). "Evidence of mechanical nonlinearity and frequency selective wave amplification in the cochlea", Archives of oto-rhino-laryngology 224(1-2):37–45. [Cited on page 18].

Kuznetsov, Yuri A. (2004). Elements of applied bifurcation theory, vol. 112 of Applied Mathematical Sciences. 3rd ed. (New York: Springer). [Cited on page 18].

Milligan, Keziah, and Nicholas Bailey (2015). "A review of software for note onset detection", in Animusic 2015. [Cited on page 16].

Moore, Brian C. J. (2012). An introduction to the psychology of hearing. (Bingley, UK: Emerald Group Publishing Limited). [Cited on page 20].

Nanavati, Sachin P., and Prasanta K. Panigrahi (2004). "Wavelet transform: A new mathematical microscope", Resonance 9:50–64. [Cited on page 17].

Olson, Elizabeth S., Hendrikus Duifhuis, and Charles R. Steele (2012). "Von békésy and cochlear mechanics", Hearing research 293(1):31–43. [Cited on page 18].

Pickles, James O. (2012). An introduction to the physiology of hearing. 4th ed. (Bingley, UK: Emerald Group Publishing Limited). [Cited on page 18].

Pullinger, Stuart (2010). A system for the analysis of musical data. Ph.D. thesis, University of Glasgow. [Cited on page 15].

Pullinger, Stuart, Nicholas Bailey, Jennifer MacRitchie, and Margaret McAllister (2009). "Computer assisted analysis and display of musical and performance data", in Proceedings of the international symposium of performance science auckland. New Zealand. [Cited on page 13].

Reetz, Henning, and Allard Jongman (2009). Phonetics: Transcription, production, acoustics, and perception. (Chichester, UK: Wiley-Blackwell). [Cited on page 4].

Ricaud, Benjamin, and Bruno Torrésani (2013). "A survey of uncertainty principles and some signal processing applications", Advances in Computational Mathematics 40(3):629–650. [Cited on page 17].

Scruton, Roger (2007). Philosophy: Principles and problems. (London: Continuum). [Cited on page 7].

Smaragdis, Paris (2009). "Relative-pitch tracking of multiple arbitrary sounds", The Journal of the Acoustical Society of America 125(5):3406–3413. [Cited on page 17].

Snow, C P (1959). The two cultures and the scientific revolution. (Cambridge University Press). [Cited on page 6].

Sundberg, Johan (1994). "Perceptual aspects of singing", Journal of Voice 8(2):106–122. [Cited on page 4].

Treitler, Leo (2011). Reflections on musical meaning and its representations, 34–35. (Bloomington & Indianapolis: Indiana University Press). [Cited on page 10].

Tymoczko, Dmitri (2011). A geometry of music: Harmony and counterpoint in the extended common practice. (Oxford/New York: Oxford University Press). [Cited on page 10].

Witten, Ian H., Eibe Frank, and Mark A. Hall (2011). Data mining: Practical machine learning tools and techniques. (Burlington, Mass.: Morgan Kaufmann). [Cited on page 16].